Introduction

$\pi(\theta)$ = prior distribution

$p(y|\theta)$ = likelihood

$\pi(\theta|y)$ = posterior distribution $\overset{\text{Bayes'}}{=} \dfrac{p(y|\theta)\pi(\theta)}{\int_\Theta p(y|\theta)\pi(\theta)\,d\theta}$ ← the inference is based on the posterior distribution

- Summaries of $\pi(\theta|y)$:
  - $\mathbb{E}[\theta|y]$
  - $Var(\theta|y)$
  - $\mathbb{P}(\theta \in C|y) \geq 0.95$ ← the interval estimate is in terms of credible intervals

- Simulate from $\pi(\theta|y)$:
  - MC
  - MCMC

- Prediction of new datapoints: $\mathcal{L}(Y_{new}|\underline{Y}=y)$ = posterior predictive distribution

$$\mathbb{P}(Y_{n+1}=1|\underline{Y}=y)? = \int_\Theta \mathbb{P}(Y_{n+1}=1|\theta)\,\pi(\theta|y)\,d\theta$$

- (Bayesian Gaussian) **Hierarchical models**: (e.g. patients in different hospitals)
  - └ two levels:
    - groups
    - units within groups

$(\underline{Y}_1, \dots, \underline{Y}_J)$ $\quad j=1,\dots,J$ groups

$\underline{Y}_j = (Y_{1,j}, \dots, Y_{n_j,j})$ $\quad i=1,\dots,n_j$ units in group j

Model:

$$Y_{1,j}, \dots, Y_{n_j,j}|\theta_j \overset{iid}{\sim} N(\theta_j, \sigma^2) \quad \text{within group}$$

$$\theta_1, \dots, \theta_J|(\mu, \tau^2) \overset{iid}{\sim} N(\mu, \tau^2) \quad \text{between-group}$$

$$(\mu, \tau^2) \sim \pi$$

With prior:

$$\frac{1}{\sigma^2} \sim gamma\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right) \quad \sigma^2 \text{ within group variance}$$

$$\frac{1}{\tau^2} \sim gamma\left(\frac{\eta_0}{2}, \frac{\eta_0\tilde{\sigma}_0^2}{2}\right) \quad \tau^2 \text{ between group variance}$$

$$\mu \sim N(\mu_0, \gamma_0^2)$$

$$\Longrightarrow \mathbb{E}[\theta_j|\bar{y}_j, \mu, \tau^2, \sigma^2] = \left(\frac{\frac{n_j}{\sigma^2}}{\frac{n_j}{\sigma^2}+\frac{1}{\tau^2}}\right)\underbrace{\bar{y}_j}_{\text{frequentist estimate of } \theta_j} + \left(\frac{\frac{1}{\tau^2}}{\frac{n_j}{\sigma^2}+\frac{1}{\tau^2}}\right)\underbrace{\mu}_{\text{prior estimate of } \theta_j}$$

Notions

- $\mathbb{P}(A|B) = \dfrac{\mathbb{P}(A,B)}{\mathbb{P}(B)}$

- $(X,Y) \sim f(x,y)$ density $\Longrightarrow$ $f_Y(y) = \int f(x,y)\,dx$, $\quad f_{X|Y=y}(x) = \dfrac{f(x,y)}{f_Y(y)}$, $\quad \mathbb{E}[X|Y] = \int_{\mathbb{R}} x \cdot f_{X|Y=y}(x)\,dx$

- $\mathbb{P}(X \in H|Y) = \mathbb{E}[\mathbb{1}_H(X)|Y]$
- $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$
- $\mathrm{Var}(X) = \mathrm{Var}(\mathbb{E}[X|Y]) + \mathbb{E}[\mathrm{Var}(X|Y)]$

$\Longrightarrow$ $\mathbb{P}(Y \leq 1 | X = \tfrac{1}{2}) = \mathbb{E}[\mathbb{1}_{(-\infty,1)}(Y) | X = \tfrac{1}{2}] = \int \mathbb{1}_{(-\infty,1)}(Y) \cdot f_{Y|X=\frac{1}{2}}(y)\,dy$

or $= \dfrac{\mathbb{P}(Y \leq 1, X = \frac{1}{2})}{\mathbb{P}(X = \frac{1}{2})}$

Bayes' theorem

Prior $\theta \sim \pi(\theta)$

Posterior $\theta | \underline{X}_n = \underline{x}$:

$$\mathbb{P}(\theta \in B | \underline{X}_n = \underline{x}) = \frac{\int_B f(\underline{x}|\theta)\,\pi(d\theta)}{\int_\Theta f(\underline{x}|\theta)\,\pi(d\theta)}$$

$f(\underline{x}|\theta)$ = density of $\underline{X}|\theta$

$$\boxed{\pi(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\,\pi(\theta)}{\int_\Theta f(\underline{x}|\theta)\,\pi(\theta)\,d\theta}}$$

marginal density of the data (we don't care too much)

if: $\begin{cases} X_1, \dots, X_n | \theta \overset{iid}{\sim} f_1(\cdot|\theta) \\ \theta \sim \pi(\theta) \end{cases}$

$$\Longrightarrow \pi(\theta|\underline{x}) = \frac{\prod_{i=1}^n f_1(x_i|\theta)\,\pi(\theta)}{\int_\Theta \prod_{i=1}^n f_1(x_i|\theta)\,\pi(\theta)\,d\theta}$$

posterior $\propto$ likelihood $\cdot$ prior

Conjugate prior: **Bernoulli–Beta** model

$X_1, \dots, X_n | \theta \overset{iid}{\sim} \mathrm{Be}(\theta)$ $\qquad \mathbb{P}(X_i = 1) = \theta$, $\mathbb{P}(X_i = 0) = 1-\theta$

$\theta \sim \mathrm{Beta}(\alpha, \beta)$

$$\longrightarrow \theta | \underline{X} = \underline{x} \sim \mathrm{Beta}\left(\alpha + \sum_{i=1}^n x_i,\ \beta + n - \sum_{i=1}^n x_i\right)$$

Beta density:

$$\pi(\theta) = \frac{1}{B(\alpha,\beta)}\,\theta^{\alpha-1}(1-\alpha)^{\beta-1}\,\mathbb{1}_{(0,1)}(\theta) \qquad \alpha, \beta > 0$$

$$\frac{1}{B(\alpha,\beta)} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}, \quad \Gamma(\alpha) = \int_0^\infty t^{\alpha-1}e^{-t}\,dt, \quad \Gamma(n) = (n-1)!, \quad \Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

Conjugate prior : **Normal** model

$$X_1,\dots,X_n \mid \mu \overset{iid}{\sim} N(\mu, \sigma_0^2) \qquad \sigma_0^2 \text{ known}$$

$$\mu \sim N(\mu_0, \tau^2) \qquad \mu_0, \tau^2 \text{ fixed}$$

$$\tau_n^2 = \frac{\sigma_0^2 \tau^2}{n\tau^2 + \sigma_0^2}$$

$$\longrightarrow \mu \mid \underline{X} = \underline{x} \sim N\left(\mu_n, \frac{\sigma_0^2}{n\tau^2+\sigma_0^2} \tau_n^2\right), \qquad \mu_n = \left(\frac{n\tau^2}{n\tau^2+\sigma_0^2}\right)\bar{x} + \left(\frac{\sigma_0^2}{n\tau^2+\sigma_0^2}\right)\mu_0$$

Conjugate prior : **Normal-Inv-Gamma** model

$$\left.\begin{aligned} X_1,\dots,X_n \mid \mu, \sigma^2 &\sim N(\mu, \sigma^2) \\ \mu \mid \sigma^2 &\sim N\left(\mu_0, \frac{\sigma^2}{n_0}\right) \\ \sigma^2 &\sim \text{inv-gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0 \sigma_0^2}{2}\right) \end{aligned}\right] := (\mu, \sigma^2) \sim \text{normal-inv-gamma}(\mu_0, n_0, \nu_0, \sigma_0^2)$$

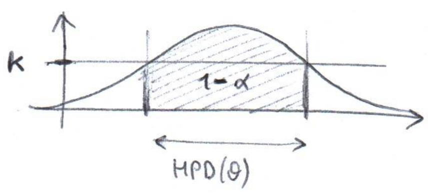
$$\Longrightarrow (\mu, \sigma^2) \sim \text{normal-inv-gamma}(\mu_1, n_1, \nu_1, \sigma_1^2)$$

$$\begin{cases} \mu_1 = \dfrac{n_0\mu_0 + n\bar{x}}{n_0 + n} \\ n_1 = n_0 + n \\ \nu_1 = \nu_0 + n \\ \sigma_1^2 = \dfrac{\frac{n_0 n}{n_0+n}(\mu - \bar{x}) + (n-1)s^2 + \nu_0\sigma_0^2}{\nu_0 + n} \end{cases} \qquad s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Inference

- Point estimation : $\left.\begin{matrix} \underline{X} \mid \theta \sim f(\underline{x} \mid \theta) \\ \theta \sim \pi(\theta) \end{matrix}\right\} \Longrightarrow \pi(\theta \mid \underline{x}) \Longrightarrow \hat{\theta}_{Bayes} = \mathbb{E}[\theta \mid \underline{x}]$

- Interval estimation : $\left.\begin{matrix} \underline{X} \mid \theta \sim f(\underline{x} \mid \theta) \\ \theta \sim \pi(\theta) \end{matrix}\right\} \Longrightarrow \pi(\theta \mid \underline{x})$
  - $C \subseteq \Theta$ is a $100 \cdot (1-\alpha)\%$ posterior credibility region if : $\mathbb{P}(\theta \in C \mid \underline{x}) \geq 1-\alpha$
  - $C \subseteq \Theta$ is a $100 \cdot (1-\alpha)\%$ posterior highest probability density region for θ if : $C = \{\theta \in \Theta : \pi(\theta \mid \underline{x}) \geq k\}$ with $k : \mathbb{P}(\theta \in K \mid \underline{x}) = 1-\alpha$



  - MCMC method: the interval estimate for θ is defined by the quantiles of the marginal posterior distribution

- Hypothesis testing : $\left.\begin{matrix} \underline{X} \mid \theta \sim f(\underline{x} \mid \theta) \\ \theta \sim \pi(\theta) \end{matrix}\right\} \Longrightarrow \pi(\theta \mid \underline{x})$

METROPOLIS - HASTINGS

We want to sample from the density $f(x)$.
We assume that we have a current value $x^{(j)}$.
We assume to have a proposal distribution $q(x \mid x^{(j)})$

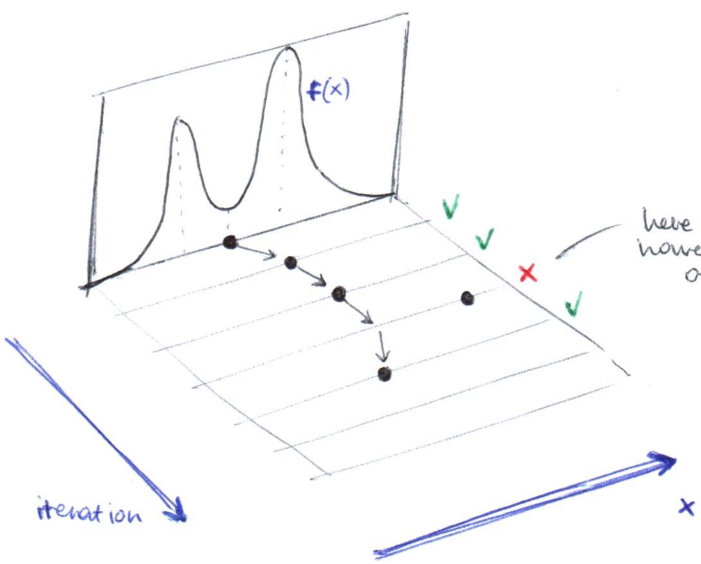
depending on the current value $x^{(j)}$

$\Rightarrow$ algorithm:

1. We sample $x^* \sim q(x \mid x^{(j)})$
2. We calculate the acceptance probability:

$$\alpha(x^{(j)}, x^*) = \min \left\{ 1, \frac{f(x^*) \; q(x^{(j)} \mid x^*)}{f(x^{(j)}) \; q(x^* \mid x^{(j)})} \right\}$$

3. Set $x^{(j+1)} = \begin{cases} x^* & \text{with probability } \alpha(x^{(j)}, x^*) \\ x^{(j)} & \text{with probability } 1 - \alpha(x^{(j)}, x^*) \end{cases}$



here we know to reject because it's really extreme, however each time we have the probability of acceptance/rejection to decide

GIBBS (mostly used in the multivariate case)

We want to sample from $f(x,y)$.
We assume to know how to sample from $f(x \mid y)$, $f(y \mid x)$.

$\Rightarrow$ algorithm:

0. Initial value $(x^{(0)}, y^{(0)})$
1. We sample $x^{(j)} \sim f(x \mid y^{(j-1)})$
2. We sample $y^{(j)} \sim f(y \mid x^{(j)})$

JAGS → gibbs sampler

STAN → based on No-U-Turn (variant of Hamiltonian Monte Carlo)